DOCUMENT RESUME

ED 036 876 24 CG 005 175

AUTHOR	Gage, N. L.
TITLE	Comments on Professor Lortie's Paper Entitled "The Cracked Cake of Educational Custom and Emerging Issues in Evaluation." Center for the Study of Evaluation of Instructional Programs Occasional Report No. 21.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-6-1646
PUB DATE	Sep 68
CONTRACT	OEC-4-6-01646-1909
NOTE	12p.; Paper from the Proceedings of the Symposium on Problems in the Evaluation of Instruction, Los Angeles, California, December 13--15, 1967
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.70
DESCRIPTORS	Curriculum Evaluation, *Educational Change, Educational Innovation, *Evaluation, *Evaluation Techniques, *Instructional Innovation, *Program Evaluation, Symposia

ABSTRACT

This commentary takes no serious exception to Professor Lortie's conclusions about educational changes and issues in evaluation, but it examines certain assumptions. It questions the implications that each school district needs independent evaluation and that evaluation should be applied at the end of a given educational program, and asks what happens when evaluation finds that a given educational innovation does not seem to make any difference. The paper offers the possibility of a more sensitive evaluation which seeks to understand more immediately realizable effects. (Author)

ED036876

BR. 6-1646
OP. 21
PA-24
CG

Occasional Report No. 21

COMMENTS ON PROFESSOR LORTIE'S PAPER ENTITLED "THE CRACKED CAKE OF EDUCATIONAL CUSTOM AND EMERGING ISSUES IN EVALUATION"

N. L. Gage

Center FOR THE
Study of Evaluation
OF INSTRUCTIONAL PROGRAMS

CG 005175

University of California, Los Angeles, September 1968

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds. Three major aspects of the program are

Instructional Variables - Research in this area will be concerned with identifying and evaluating the effects of instructional variables, and with the development of conceptual models, learning theory and theory of instruction. The research involves the experimental study of the effects of differences in instruction as they may interact with individual differences among students.

Contextual Variables - Research in this area will be concerned with measuring and evaluating differences in community and school environments and the interactions of both with instructional programs. It will also involve evaluating variations in student and teacher characteristics and administrative organization.

Criterion Measures - Research in this field is concerned with creating a new conceptualization of evaluation of instruction and in developing new instruments to evaluate knowledge acquired in school by measuring observable changes in cognitive, affective and physiological behavior. It will also involve evaluating the cost-effectiveness of instructional programs.

ED036876

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

COMMENTS ON PROFESSOR LORTIE'S PAPER ENTITLED "THE CRACKED CAKE OF EDUCATIONAL CUSTOM AND EMERGING ISSUES IN EVALUATION"

N. L. Gage

Stanford University

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation of Instructional Programs

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSEIP Occasional Report No. 21, September 1968
University of California, Los Angeles

COMMENTS ON PROFESSOR LORTIE'S PAPER ENTITLED "THE CRACKED CAKE OF EDUCATIONAL CUSTOM AND EMERGING ISSUES IN EVALUATION"

N. L. Gage

Professor Lortie has presented an incisive analysis of the forces operating to make evaluation a central concern of contemporary educational enterprise. Applying the conceptual tools of the sociologist, he has furnished what might be called an apologia pro centro UCLA. That is, one leaves his paper with the feeling that whatever else may occur in American education in the years ahead, a major part of it will consist of evaluation.

An Overview of Lortie's Paper

External trends resulting from federal activities, business corporations, and universities are increasing the pressure to evaluate because of the many new alternatives that they are placing before school personnel. And the trends internal to the school system, such as the specialization and stratification of educational workers and the increase in teacher militancy, will also enhance the need to evaluate because the rivalries among different categories of educational workers will lead them to attempt to strengthen their positions by means of evaluative data.

Further, the increased rate of change in American education, resulting in part from the development of substitutes for the old participatory model of school administration, has brought about greater reliance upon evaluative efforts by specialists in such work. Similarly, the enlarged scope of change in education, which

affects previously untouched aspects of student life, calls for increased emphasis on evaluation. Team teaching, nongraded classrooms, and the particular social forms of a given kind of professional training all tend correspondingly to broaden the scope of the variables with which evaluation workers must be concerned.

Dr. Lortie's third section deals with the effects on evaluation work of the increasing recognition of education as a system of forces affecting every aspect of society--economic, social, political, and cultural. Education, in turn, is influenced by a similarly broad array of forces, ranging from explosions in population and knowledge to the computer revolution and the labor market. The kinds of concerns important to evaluators are similarly enlarged; it will no longer suffice to concentrate merely on what goes on in schools and classrooms if evaluation is to have a correspondingly broadened scope.

Next, Dr. Lortie asks where the increased amount of evaluative talent is to orginate, how greater knowledge and understanding of evaluative procedures can be disseminated, and how more specialists can be produced and deployed. Second, evaluation workers may need to concern themselves with decision-making processes in education and their relationship to the feedback of evaluative data on proposed changes. Third, he sees the problem of maintaining public confidence in the integrity and validity of educational evaluations made by outside experts. Models of such efforts in medicine, accounting, and other fields suggest the emergence of a fee-for-service profession of autonomous evaluators. Further, Dr. Lortie pointed to the issues of moral complexity

involved in evaluation against any set of objectives to be attained by a set of means. For many of these means-end combinations we do not know enough about their latent functions to have confidence in our value judgments.

Finally, Lortie raises the question of whether many of the kinds of expertise needed in adequate evaluation must be drawn from other domains, such as economics, crime prevention, or race relations. If so, the implications for specialization and training of evaluation workers must be given thought.

A Discussion

I can find no reason to take serious exception to most of what Dr. Lortie has offered. But there are some aspects of his analysis that seem to rest on assumptions that ought to be made explicit and perhaps be subjected to some questioning.

1. It seems he implies that every school district--or, indeed, every school--will need to evaluate independently all innovations or educational alternatives that may be developed and offered in the years ahead. It is almost as if he saw the need for each school board to operate its own consumer's union for the evaluation of new developments, procedures, and products offered to the schools.

It is unclear to me whether he predicts or supports any such conception of the role of evaluation, which, in either case, ought to be severely questioned. Surely, the differences between our school districts are not so great in any of the relevant dimensions that they require each district to make its own independent evaluation of proposed innovations in the schools. It must be possible

for any given school district to learn something from the experience of others, that is, from the reports of large-scale evaluative efforts conducted in representative samples of school districts and classrooms. Just as every prospective car buyer need not be an expert in compression ratios and crankshaft bearings, so every school board need not arrange for its own evaluation of team teaching or the nongraded classroom.

Perhaps a distinction ought to be made between the evaluation of an innovation in its general form and the evaluation of how it is working in a particular and local situation. Such a distinction would be analogous to that between the kind of evaluation made by Consumer's Union, concerning how good the 1968 Fairlane is in general, and the evaluation made by the owner and operator of a particular Fairlane concerning how well his own car is working. For the former kind of evaluation, large-scale programs of evaluation by experts ought to do the job for 10,000 school districts at a time; for the latter kind of evaluation, each school district, school, administrator, or teacher will need to know how to tell whether a given practice, old or new, is working well and when something is seriously wrong. It is the difference between the kind of evaluation made by the automotive engineer and that made by you and me, who know enough to get worried when the exhaust from our car gets too black.

2. Another question can be raised concerning Dr. Lortie's implicit conception of evaluation. If I understood the connotations of his discussion, he sees evaluation as consisting of a kind of posttesting applied at the end of the operation of a given

kind of educational innovation or program. But many students of evaluation have questioned this sort of conception. They say that evaluation merely at the end of a curriculum development program is often ineffectual. Evaluative effort ought to be poured into educational innovations while they are being developed, as part of the developmental process itself. The evaluation worker ought to work with the educational innovator, the developer of a new kind of evaluational procedure or material. In this sense, evaluation takes place much more frequently on the basis of a much wider variety of evidence. The analogy that occurs to me here is the difference between the kind of evaluation of a student's learning that takes place after each frame in a body of programmed instruction and the kind of evaluation that takes the form of a final examination at the end of a course. If I understand him, Dr. Lortie deals only with the latter kind of evaluation. It would have been desirable if his analytic effort had also been turned to the former.

3. Next, I should like to raise the question of what happens to evaluation when it finds, time after time, that a given kind of educational innovation or alternative does not seem to make any difference. J. M. Stephens, in his recent little book, The Process of Schooling, has pointed out that most evaluations of educational innovations have yielded negative results. Stephens documented his position with references to summaries of studies of a host of specific educational variables, procedures, practices, and orientations. That is, he summarized the summaries of studies of school attendance, instructional television, independent study and correspondence courses, size of class, individual consultation

and tutoring, counseling, concentration on specific students, the students' involvement, the amount of time spent in study, distraction by jobs and extracurricular activities, size of school, the qualities of teachers that can be rated by principals and supervisors, nongraded schools, team teaching, ability grouping, progressivism versus traditionalism, discussion versus lecture, group-centered versus teacher-centered approaches, the use of frequent quizzes, and programmed instruction. According to Stephens, studies of all these have failed to show that they make a consistent and significant difference.

Stephens briefly considered the possibility that the negative results are due to methodological errors such as concentration on one narrow segment of achievement, using insensitive tests, employing poor controls, exerting overcontrol that holds too much constant and so restricts the differences, or using a too stringent criterion of statistical significance. Stephens concluded that negative results are only to be expected, because "in the typical comparison of two administrative devices (such as teaching methods) we have two groups that are comparable in the forces responsible for (say) 95 percent of the growth to be had and which differ only in the force that, at best, can affect only a small fraction of the growth" (Stephens, 1967, p. 84).

This is not the place for any extended discussion of the exact merits of the details of Stephen's argument. But I cannot disagree with his statement that, by and large, evaluations of innovative efforts in education have yielded a "flood of negative results." Instances come to mind almost every day.

Recently, for example, the "More Effective Schools" program in the New York City schools was found by the Center for Urban Education to make no significant difference in the achievement of the pupils in the participating schools. Evaluations of Project Head Start have been more noteworthy for their failure to reveal significant differences than for anything else.

If so, what are the implications for evaluation workers? Are they willing to be cast in the role of the spoilsport, who continually finds nothing in favor of, as well as nothing much against, the shiniest products of the ingenious innovator's art? Will evaluation come to be regarded, even more than it is now, as a threat to the fondest hopes of the educational thinker, research worker, and developer? Will mechanisms be developed to rationalize, even more than is now the case, the unwillingness of the educational worker to subject his pet scheme to the kinds of tests that, nowadays at least, 95 times out of 100, reveal no statistically significant difference, one way or another? If as Dr. Lortie said, there is "no tradition of tough-minded empirical evaluation among American teachers and administrators" (p. 18), perhaps the reason is that they have so often in the past gone without reinforcement when they have exhibited evaluative behavior.

But another question also seems to arise almost inevitably. One wonders whether the failure to find significant results when educational innovations are evaluated stems not from some weakness of the innovations, as Stephens thinks, but rather from some weakness in the kinds of evaluations that are made. Perhaps the gross kind of measure of the effect of a given innovation is,

indeed, too insensitive to reveal the effects of that treatment. Perhaps we need to become much more modest in the kinds of effects we seek. Perhaps we need to bring them much closer in time, space, and conceptualization to the operation of the kinds of educational variables that are manipulated in our experiments.

My analogy here is the difference between dropping a rock on the surface of a lake, then seeking its effect within 20 feet, and seekings its effect a mile away. In principle the rock does affect the water a mile away, but the effect has become much too attenuated at that distance to be discernible. Similarly, the effects of team teaching or the nongraded classroom or programmed instruction ought not to be sought at the end of the school year in terms of the mean score differences between an experimental and a control group; rather, they ought to be sought in the day-to-day behaviors of teachers and pupils--an approach much closer in time, space, and conceptual relevance to the kind of innovation being studied. Someday, understanding of these more modest kinds of effects will culminate in an understanding of the chain of effects with long-range and large-scale significance of the kind that is nowadays more often evaluated. But in the meantime it might make better sense--and lead to less frustration caused by negative results--if evaluators focused more sharply on variables that have some chance of being affected by the kinds of innovations they study.

Dr. Lortie paid no attention, as I understood him, to the role of evaluation in improving scientific understanding of educational processes as separate from the role of evaluation affecting

practical decision-making in the schools. Yet the concerns of the former kind of evaluation may in the long run be more important than the latter. Such evaluation for the purposes of scientific research can lead to greater understanding of pupils and better theory about what occurs in classrooms, schools, and school districts. It is surely unnecessary to defend, before this audience, the values for educational practice of improved theoretical formulations.